

基于多层次项目相似度的协同过滤推荐算法

徐翔宇 刘建明

(桂林电子科技大学计算机科学与工程学院 桂林 541004)

摘 要 针对传统的基于项目的协同过滤推荐算法中项目相似度的计算上存在的缺陷,提出一种基于多层次项目相似度的协同过滤推荐(MLCF)算法。利用多维度启发式方法分析用户行为记录,从共同用户集、用户活跃度、项目得分时效和项目得分4个方面综合分析项目之间的相似程度,并在此基础上,设计多层次项目相似度计算方法。实验结果表明,基于多层次项目相似度的推荐算法相对于传统的基于项目的协同过滤推荐算法具有较高的推荐准确率、召回率和较低的平均绝对误差值。

关键词 协同过滤,启发式方法,多层次,项目相似度

中图分类号 TP301 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.10.049

Collaborative Filtering Recommendation Algorithm Based on Multi-level Item Similarity

XU Xiang-yu LIU Jian-ming

(School of Computer Science and Engineering, Guilin University of Electronic Technology, Guilin 541004, China)

Abstract For the defects in the calculation of item similarity of traditional item-based collaborative filtering, this paper proposed an improved collaborative filtering algorithm based on multi-level item similarity. Firstly, the multi-dimensional heuristic methods are used to analyze the similarity of items comprehensively by analyzing user's behavior records in four aspects, including user collective rating items, user activity, user rating timeliness and user rating. Secondly, based on the four aspects of item similarity, a method for calculating multi-level item similarity is designed. Experimental results show that, compared with the traditional item-based collaborative filtering recommendation algorithm, the algorithm based on multi-level item similarity has higher recommendation accuracy rate and recall rate, and lower MAE value.

Keywords Collaborative filtering, Heuristic methods, Multilevel, Item similarity

1 概述

互联网信息的疯狂增长,为用户获取信息提供了便利,但同时也使得用户探寻有价值的信息变得更加困难,导致用户查找有用信息的效率下降,这就是海量信息导致的信息超载(information overload)的问题^[1]。搜索引擎能帮助用户找到对应的信息,但这种方式需要用户提供关键字,然而多数情况下,用户无法提供明确的需求,因此急需一个自动化工具通过对用户历史行为记录进行分析,从海量信息中找到用户感兴趣的信息来供用户选择。该自动化工具就是推荐系统。目前推荐系统有电子商务、电影、音乐电台、新闻、社交平台等多个应用领域,对应的典型公司有 Amazon、Netflix、Pandora、Yahoo、Facebook 等。

在多种推荐算法中,协同过滤算法得到了非常广泛的应用。该算法可以分为基于用户的协同过滤算法(user-based collaborative)和基于项目的协同过滤算法^[2](item-based col-

laborative filtering)。其中,前者通过分析用户的历史数据来计算用户之间的相似度,然后依靠近邻用户来提供推荐服务。后者通过分析用户作用于项目的行为数据,计算项目之间的相似度,根据项目相似度与用户的历史兴趣为用户进行推荐。余弦相似度^[3]、Pearson 相关系数和修正的余弦相似度等都属于传统的相似度计算方法。可以看出,相似度的计算是推荐系统的核心所在,如何挖掘相似用户或相似项目一直是该算法领域中研究的重点。例如,文献[4]针对传统的相似度量方法在计算最近邻上的缺陷,引入项目评分预测,提高了推荐精度。文献[5]针对单一评分相似性计算导致的推荐质量不高的问题,将项目分类、评分值及次数引入特征向量,计算同组项目之间的相似性,提出了一种基于项目兴趣度的协同过滤算法,提高了推荐质量。文献[6]通过 Tag 分类信息,得到项目的最近邻,对用户的评分值进行预测,填充用户评分矩阵,使得推荐系统的平均绝对误差得到降低。文献[7]针对现有推荐算法对项目类别的忽视,提出了把项目类别相似性融入

到稿日期:2015-07-30 返修日期:2016-02-03 本文受国家自然科学基金(61262074),广西可信软件重点实验室开放课题(kx201101),广西高校优秀人才资助计划(桂教人 201065),广西自然科学基金(2012GXNSFCA053009),广西信息科学实验中心项目,桂林电子科技大学计算机软件创新团队资助。

徐翔宇(1989—),男,硕士生,主要研究方向为推荐系统;刘建明(1975—),男,博士后,教授,博士生导师,主要研究方向为计算机及通信网络组网工程。

评分相似性并进行加权的方法,提高了查找最近邻居项目的准确度。文献[8]通过对用户历史记录的分析,从兴趣广度、深度和时效3个方面分析用户兴趣,设计了三级计算方法来进一步挖掘相似用户,提高了个性化推荐的准确率。

上述研究^[4-7]在寻找相似项目的过程中,虽然从不同方面提高了挖掘相似项目的准确度,但是仍然未更深入地分析多个维度对项目相似性的影响,对项目相似度的计算大都忽视了活跃度和项目得分时效性对项目相似度的影响,从而没有进一步提高挖掘相似项目的能力,影响了推荐效果。文献[8]从3个维度设计了用户之间兴趣度的三级计算方法,但是并没有考虑用户活跃度对用户之间相似度造成的影响,并且上网用户数量的飞速增长将会导致在计算用户之间相似度的时间复杂度和空间复杂度将会与用户数量的增长呈近似于平方关系,相比项目相似度的计算,计算量会更加庞大。

针对这些问题,本文利用多维度启发式方法分析用户行为记录,在此基础上设计多层次项目相似度计算方法,提出基于多层次项目相似度计算的协同过滤推荐算法,进一步挖掘相似项目,从而提高推荐准确率。

2 多维度启发式方法分析用户行为记录

2.1 用户行为记录形式化表示

传统的基于项目的协同过滤算法中,单一利用评分矩阵分析项目之间相似度比较片面。本文提出的多层次项目相似度计算主要从用户对项目的评分矩阵和时间矩阵中挖掘相似项目。从评分矩阵中,除了利用项目共同用户集和项目得分对相似项目进行挖掘外,还进一步分析了用户活跃度对相似项目的影响;通过时间矩阵,深入分析了项目得分时效对相似项目的影响。通过对两个矩阵的多维度分析,全方位描述相似项目,进而更准确地挖掘出相似项目。

本文主要依据用户行为中项目的得分和得分时间记录来挖掘相似项目。通常情况下,用户的此类行为记录中包括用户 u 的 ID_u 、项目 i 、用户 u 对项目 i 的评分 S_{ui} 、用户 u 对项目 i 的评分时间 T_{ui} ,所以用户 u 对项目 i 的行为记录可以形式化定义为:

$$\langle Behavior(u,i) \rangle ::= \langle ID_u \rangle \langle i \rangle \langle S_{ui} \rangle \langle T_{ui} \rangle$$

2.2 多维度启发式方法描述项目相似性

项目是否相似不能简单地查看用户对项目的评分,还应联系用户对项目的更多行为记录来进行综合分析。认知心理学的信息加工理论认为^[9],启发式是人类思维解决问题的重要方法。根据启发式的方法,可以得到用户行为记录可能对项目相似度产生影响的4个启发:

- (1)被相同用户选择的项目之间相似度相对较高;
- (2)活跃度大的用户对项目相似度的贡献应该小于活跃度小的用户;
- (3)相同用户对不同项目产生行为的时间越接近,项目相似度越高;
- (4)被相同用户选择的项目之间得分越接近,项目相似度越高。

基于上述4个启发,本文提出基于多个维度的启发式规则,并在之后利用这些启发式规则设计多层次项目相似度计

算方法。本文中记项目 i 与 j 之间的相似度为 $Sim(i,j)$ 。

2.2.1 基于共同用户集的启发式规则

首先从项目共同用户集这个维度对项目相似度进行分析。在用户对项目的行为记录 $Behavior(u,i)$ 中,可以将作用于每个项目 i 的用户集统计出来,记为 $U_{srSet}(i) = \{user1, user2, \dots\}$ 。

直观地,项目 i 和 j 的共同用户集中交集元素越多,代表两个项目同时属于越多相同用户的兴趣范围之内,项目 i,j 的相似度就越高,在选择项目 i 的相似项目时,项目 j 就更适合被考虑到。

从上述启发中,可以定义基于共同用户集的启发式规则:设作用于项目 i,j,k 的用户集分别为 $U_{srSet}(i), U_{srSet}(j), U_{srSet}(k)$,如果 $|U_{srSet}(i) \cap U_{srSet}(j)| > |U_{srSet}(i) \cap U_{srSet}(k)|$,那么项目之间的相似度 $Sim(i,j) > Sim(i,k)$ 。

2.2.2 基于活跃度的启发式规则

在用户对项目的行为记录 $Behavior(u,i)$ 中,可以利用用户 u 的评分项目数来衡量用户 u 的活跃度,将用户 u 的评分项目数记为 $RatingNum(u)$ 。

直观地,活跃度越高的用户作用的项目对项目相似度计算上的贡献越低。例如,某用户由于开设音像店,购买了亚马逊网站70%以上的唱片用于个人音像店销售,这意味着该用户的行为记录会使得这70%以上的唱片之间产生相似度,然而这位活跃的用户购买的这些唱片并非出于自身的兴趣,所以他所购买的唱片之间的相似度贡献值应该远小于一个只买了几张披头士唱片的摇滚音乐爱好者。

从上述启发中,可以定义基于活跃度的启发式规则:

设用户 u 和用户 v 的评分项数分别为 $RatingNum(u), RatingNum(v)$,如果 $|RatingNum(u)| > |RatingNum(v)|$,那么利用用户 v 的行为记录计算出的项目 i,j 之间的相似度应该大于利用用户 v 的行为记录计算出的项目 i,k 之间的相似度,即 $Sim(i,j) > Sim(i,k)$ 。

2.2.3 基于时效的启发式规则

在用户对项目的行为记录 $Behavior(u,i)$ 中,通过用户-项目作用时间矩阵,可以将项目 i 受用户作用时间的集合统计出来,记为 $TimeSet(i) = \{time1, time2, \dots\}$ 。

直观地,用户 u 对项目 i 和项目 j 产生行为的时间间隔越近,项目 i 和项目 j 的相似度就越高。例如用户 u 在同一天观看了《八公犬的故事》和《忠狗德拉姆》两部电影,随着用户兴趣的漂移^[10],一个月后该用户观看了电影《深入敌后》,那么从时间间隔上来说,同一天被观看的两部电影《八公犬的故事》和《忠狗德拉姆》之间的相似度要大于电影《八公犬的故事》和《深入敌后》之间的相似度。

从上述启发中,可以定义基于时效的启发式规则:

设用户 u 对项目 i,j,k 的作用时间分别为 $Time_{ui}, Time_{uj}, Time_{uk}$,如果 $|Time_{ui} - Time_{uj}| > |Time_{ui} - Time_{uj}|$,那么项目 i,j 之间的相似度应该大于项目 i,k 之间的相似度,即 $Sim(i,j) > Sim(i,k)$ 。

2.2.4 基于项目得分的启发式规则

在用户对项目的行为记录 $Behavior(u,i)$ 中,利用用户-项目评分矩阵,可以将项目 i 被不同用户评分集合统计出来,

记为 $ScoreSet(i) = \{score1, score2, \dots\}$ 。

直观地,用户 u 对项目 i 和项目 j 的评分越接近,则项目 i 和项目 j 的相似度就越高。在选择与项目 i 相似的项目时,项目 j 就更适合被考虑到。

从上述启发中,可以定义基于项目得分的启发式规则:

设用户 u 对项目 i, j, k 的评分分别为 $Score_{ui}, Score_{uj}, Score_{uk}$, 如果 $|Score_{ui} - Score_{uj}| > |Score_{ui} - Score_{uk}|$, 那么项目 i, j 之间的相似度应该大于项目 i, k 之间的相似度, 即 $Sim(i, j) > Sim(i, k)$ 。

3 项目相似度的多级计算方法

对于项目之间共同用户集的相似性,本文沿用 Jacard 公式来进行计算,在随后的逐级计算中,基于上文中提出的多维度启发式规则,逐步融入其它维度因素,对项目的相似度进行改进。

3.1 基于共同用户集的第1级相似度计算

设项目 i, j 的用户集分别为 $U_{sr}(i), U_{sr}(j)$, 根据多维度启发式方法中基于共同用户集的启发式规则,通过 Jacard 公式来计算基于共同项目集的相似度,可以将项目 i 和项目 j 的相似度定义为:

$$Sim^{(1)}(i, j) = \frac{|U_{sr}(i) \cap U_{sr}(j)|}{\sqrt{|U_{sr}(i)| \cdot |U_{sr}(j)|}} \quad (1)$$

从式(1)可以看出,项目 i 和项目 j 之间共同用户集交集越大,项目 i, j 之间的相似度就越高,此公式沿用了 Jacard 公式,符合上文中提出的基于共同用户集的启发式规则思想。

3.2 融入活跃度的第2级相似度计算

设项目 i 和项目 j 的共同用户集为 $ComU_{sr}(i, j) = U_{sr}(i) \cap U_{sr}(j)$, 根据多维度启发式方法中基于活跃度的启发式规则,可以将用户 u 的活跃度通过用户 u 的评分项目数量 $RatingNum(u)$ 来表示,进而为项目 i 和 j 之间的相似度计算融入活跃度影响,改进为第2级相似度计算,表达形式如下:

$$Sim^{(2)}(i, j) = Sim^{(1)}(i, j) \cdot \frac{\sum_{u \in ComU_{sr}(i, j)} 1}{\ln(1 + |RatingNum(u)|)} \quad (2)$$

从式(2)可以看出,活跃度越高的用户,对项目之间相似度的贡献越小,这符合上文中提出的基于活跃度的启发式规则思想。

3.3 融入时效的第3级相似度计算

设用户 u 对项目 i 和项目 j 的评分时间分别为 $Time_{ui}$ 和 $Time_{uj}$, 共同用户集仍为 $ComU_{sr}(i, j)$, 根据多维度启发式方法中基于时效的启发式规则,可以为项目 i 和项目 j 之间的相似度计算融入时间差值的影响,改进为第3级相似度计算,表达形式如下:

$$Sim^{(3)}(i, j) = Sim^{(2)}(i, j) \cdot \frac{\sum_{u \in ComU_{sr}(i, j)} 1}{1 + |Time_{ui} - Time_{uj}|} \quad (3)$$

从式(3)可以看出,相同用户对项目 i, j 评分时间差越小,项目 i, j 之间的相似度就越大,这符合上文中提出的基于时效的启发式规则思想。

3.4 融入项目得分的第4级相似度计算

设用户 u 对项目 i 和项目 j 的评分分别为 $Score_{ui}$ 和

$Score_{uj}$, 共同用户集仍为 $ComU_{sr}(i, j)$, 根据多维度启发式方法中基于项目得分的启发式规则,可以为项目 i 和项目 j 之间的相似度计算融入项目得分差值影响,改进为第4级相似度计算,表达形式如下:

$$Sim^{(4)}(i, j) = Sim^{(3)}(i, j) \cdot \frac{\sum_{u \in ComU_{sr}(i, j)} 1}{1 + |Score_{ui} - Score_{uj}|} \quad (4)$$

从式(4)可以看出,相同用户对项目 i, j 评分分值差越小,项目 i, j 之间的相似度就越大,这符合上文中提出的基于项目得分的启发式规则思想。

以上给出的多级相似度计算能够从不同的角度分析项目之间的相似性,从而提高推荐效果。这将在本文的实验部分得到验证。

4 基于多层次项目相似度计算的协同过滤推荐

基于项目的协同过滤算法的主要步骤为:1)计算项目之间的相似度,据此得到每个项目的最近邻项目集;2)根据项目的相似度和用户的历史兴趣为用户生成推荐列表。

4.1 用户对项目感兴趣程度的计算

上文基于多维度启发式规则得出了项目之间相似度的多级计算方法,因此可以将用户 u 对未产生行为的项目 j 的兴趣度表示为:

$$p_{uj} = \sum_{i \in I(u) \cap S(j, K)} Sim^{(4)}(j, i) \times R_{ui} \quad (5)$$

其中, $I(u)$ 表示用户 u 感兴趣的项目集合, $S(j, K)$ 表示与项目 j 最相似的 K 个项目的集合,可以通过对 $Sim^{(4)}(j, i)$ 进行 Top- N 运算求得, R_{ui} 为用户 u 对项目 i 的感兴趣程度,在本文中通过评分来体现。由式(5)可知,根据项目之间的相似度和用户历史兴趣,可得到用户对未产生项目的兴趣程度,这符合基于项目的协同过滤算法思想。

4.2 算法步骤

如上文所述,基于多级项目相似度计算的协同过滤算法可以分为两个步骤:

(1)通过式(4)和 Top- N 运算求得项目 j 的最近邻项目集合 $S(j, K)$;

(2)通过式(5)计算用户 u 对项目 j 的感兴趣程度,向用户 u 推荐 p_{uj} 中 Top(K)的项目集合。

依据这两个步骤,可以对算法进行详细描述。

步骤1 求项目 j 的最近邻项目集合 $S(j, K)$

输入:项目 j , 全部项目集合 Items

输出:项目 j 的最近邻集合 $S(j, K)$

Begin:

```
for i in Items{
    if(U_{sr}(j) \cap U_{sr}(i) \neq NULL){
        Comusr = U_{sr}(i) \cap U_{sr}(j);
        Sim^{(1)}(j, i) = |U_{sr}(i) \cap U_{sr}(j)| / \sqrt{|U_{sr}(i)| \cdot |U_{sr}(j)|};
        Acty = 0; Time = 0; Score = 0;
        for u in Comusr{
            Acty += 1 / [\ln(1 + RatingNum(u))];
            Time += 1 / (1 + |Time_{ui} - Time_{uj}|);
            Score += 1 / (1 + |Score_{ui} - Score_{uj}|);
        }
        Sim^{(2)}(i, j) = Sim^{(1)}(j, i) \cdot Acty;
```

$$\text{Sim}^{(3)}(i,j)=\text{Sim}^{(2)}(j,i) \cdot \text{Time};$$

$$\text{Sim}^{(4)}(i,j)=\text{Sim}^{(3)}(j,i) \cdot \text{Score}; \}$$

for i in Items

TopN($\text{Sim}^{(4)}(i,j)$); //得到项目 j 最近邻集合 $S(j,k)$

End

该过程的时间复杂度为 $O(mn)+O(mlgm)$, 其中, m 为全部项目数, 即 $m=|\text{Items}|$, n 为对项目 i 与其它项目的共同用户的最大个数, 即 $n=\text{Max}(\text{Comusr})$ 。时间复杂度的第一项 $O(mn)$ 是计算项目 j 和其它所有项目相似度的时间复杂度, 由于 $n \ll m$, 因此 $O(mn)$ 可以简化为 $O(m)$, 时间复杂度的第二项 $O(mlgm)$ 为堆排序 $\text{Sim}^{(4)}(j,i)$ 的时间复杂度。综上, 该过程的时间复杂度最终可确定为 $O(mlgm)$ 。

步骤 2 求出用户 u 对未产生行为的项目 j 感兴趣程度 $\text{Top}(K)$ 的项目集合

输入: 项目 j 的最近邻项目集合 $S(j,K)$, 用户 u 未产生行为的项目集
输出: 用户 u 对未产生行为的项目 j 感兴趣程度 $\text{Top}(K)$ 的项目集合
Begin:

for j in Items and not in List(u) {

$R(u,j)=0$;

for i in $S(j,K)$ {

if ($i \in \text{List}(u)$)

$R(u,j)=R(u,j)+\text{Sim}^{(4)}(i,j) \cdot \text{Score}_{ui}$;

for j in Items

TopN($R(u,j)$); //得到为用户 u 推荐的项目集合

End

该过程的时间复杂度为 $O(mk)+O(mlgm)$, 其中 m 是用户 u 未产生行为的项目数量, k 为最近邻项目数, 时间复杂度第一项 $O(mk)$ 是计算用户 u 对其未产生行为的项目 j 感兴趣程度的时间复杂度, 由于 $k \ll m$, 因此 $O(mk)$ 可以简化为 $O(m)$, 时间复杂度的第二项 $O(mlgm)$ 为堆排序 $R(u,j)$ 的时间复杂度。综上, 该过程的时间复杂度最终可确定为 $O(mlgm)$ 。

5 实验结果与分析

5.1 实验数据

本文使用的是被应用于音乐推荐领域的 KDD Cup 2011 数据集^[11], 该数据集包含的信息有: 音乐爱好者 ID(userID)、音乐 ID(ItemID)、用户对听过的音乐的评分(Rating)、用户听音乐的时间(Timestamp)。所以该数据集能够提供本文阐述的用户 u 对项目 i 的行为记录形式化定义中所需的信息, 即 $\langle \text{Behavior}(u,i) \rangle ::= \langle \text{ID}_u \rangle \langle i \rangle \langle S_{ui} \rangle \langle T_{ui} \rangle$ 。整个数据集将被划分为训练集和测试集, 其中训练集将被用于评分预测并进行推荐, 测试集将被用来测试算法性能。

5.2 评价指标

在进行 Top-N 推荐时, 会为用户提供一个推荐列表, 所以本文采用的评价指标为推荐的准确率(Precision)、召回率^[12](Recall)和平均绝对误差值(MAE)。

准确率定义为:

$$\text{Precision} = \frac{\sum_{u \in U_{sr}} |R(u) \cap T(u)|}{\sum_{u \in U_{sr}} |R(u)|} \quad (6)$$

召回率定义为:

$$\text{Recall} = \frac{\sum_{u \in U_{sr}} |R(u) \cap T(u)|}{\sum_{u \in U_{sr}} |T(u)|} \quad (7)$$

平均绝对误差值定义为:

$$\text{MAE} = \frac{\sum_{u,i \in T} |r_{ui} - \hat{r}_{ui}|}{|T|} \quad (8)$$

其中, $R(u)$ 是推荐系统根据用户在训练集上的行为, 给用户提供的推荐列表; $T(u)$ 是用户在测试集上的行为列表。通过定义可以看出, 准确率定义了推荐列表中包含的用户行为占用户在测试集上行为记录的比例, 召回率定义了推荐列表中包含的用户行为占推荐列表中项目数的比例, MAE 定义了预测评分与用户实际评分之间的误差度量。

5.3 结果分析

实验中, 将本文提出的基于多层次项目相似度计算的算法与传统的基于项目推荐算法^[2]和改进的推荐算法^[4]进行比较, 推荐的列表长度依次选为 10, 15, 20, 25, 30。运行出各推荐算法的准确率、召回率和 MAE 数值, 同时生成折线图进行对比。实验结果如图 1—图 3 所示。

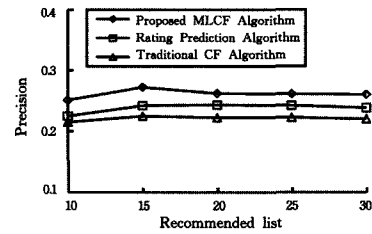


图 1 准确率实验结果

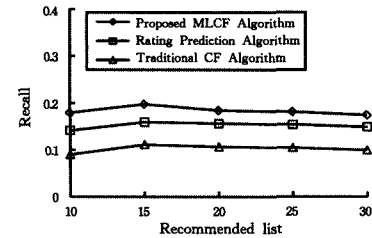


图 2 召回率实验结果

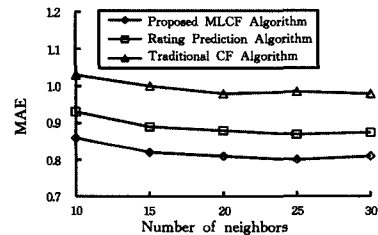


图 3 MAE 实验结果

通过准确率、召回率和 MAE 值的实验结果可以看出, 利用多层次项目相似度计算的方法在基于项目的协同过滤算法中推荐结果的准确率和召回率都有提高, 且具有较低的 MAE 值。

结束语 由于信息的飞速增长, 推荐系统在很多领域得到了应用。面向基于项目的协同过滤推荐领域, 本文提出了基于多层次项目相似度计算的协同过滤推荐算法。该算法主要是通过利用多维度启发式方法分析用户的行为记录, 从多

(下转第 291 页)

- [3] Zhang S Q, Chen L, Xiong W. Research on performances of parallel programming models based on chip multi-processor[C]// 2012 IEEE International Conference on Oxide Materials for Electronic Engineering (OMEE). 2012; 632-635
- [4] Wang S. CyberGIS: Blueprint for integrated and scalable geospatial software ecosystems[J]. International Journal of Geographical Information Science, 2013, 27(11): 2119-2121
- [5] Porwal S, Katiyar S K. Performance evaluation of various resampling techniques on IRS imagery[C]// 2014 Seventh International Conference on Contemporary Computing (IC3). IEEE, 2014; 489-494
- [6] Cai Zi-xing, Peng Meng, Yu Ling-li. Adaptive Incremental Principal Component Analysis Visual Tracking Method Based on Temporal Characteristics[J]. JEIT, 2015, 37(11): 2571-2577 (in Chinese)
蔡自兴, 彭梦, 余伶俐. 基于时序特性的自适应增量主成分分析的视觉跟踪[J]. 电子与信息学报, 2015, 37(11): 2571-2577
- [7] Xu Miao-zhong, Cong Ming, Wan Li-juan, et al. A Methodology of Image Segmentation for High Resolution Remote Sensing Image Based on Visual System and Markov Random Field[J]. Acta Geodaetica et Cartographica Sinica, 2015, 44(2): 198-205 (in Chinese)
- 许妙忠, 丛铭, 万丽娟, 等. 视觉感受与 Markov 随机场相结合的高分辨率遥感影像分割法[J]. 测绘学报, 2015, 44(2): 198-205
- [8] Zhang Wei, Xiong Wei-li, Xu Bao-guo. Multi-model Combination Modeling Based on Just-in-time Learning Using Gaussian Process Regression[J]. Information and Control, 2015, 44(4): 487-492, 498 (in Chinese)
张伟, 熊伟丽, 徐保国. 基于实时学习的高斯过程回归多模型融合建模[J]. 信息与控制, 2015, 44(4): 487-492, 498
- [9] Rasmussen C E, Williams C K I. Gaussian processes for machine learning[M]. Cambridge, MA, US: The MIT Press, 2006
- [10] Jiang Jian-guo, Jin Yu-long, Qi Mei-bin, et al. Moving Target Detection in Natural Scene Based on Sparse Representation of Residuals[J]. Chinese Journal of Electronics, 2015, 43(9): 1738-1744 (in Chinese)
蒋建国, 金玉龙, 齐美彬, 等. 基于稀疏表达残差的自然场景运动目标检测[J]. 电子学报, 2015, 43(9): 1738-1744
- [11] Zhou Yu-gu, Wang Ping, Gao Yin-hui. Remote Sensing Image Classification with Bag-of-visual-words Model[J]. Journal of Chongqing University of Technology (Natural Science), 2015, 29(5): 71-77 (in Chinese)
周宇谷, 王平, 高颖慧. 基于视觉词袋模型的遥感图像分类方法[J]. 重庆理工大学学报(自然科学), 2015, 29(5): 71-77

(上接第 265 页)

个维度挖掘相似项目。实验结果证明, 与传统的基于项目的协同过滤算法相比, 本文提出的基于多层次项目相似度计算的协同过滤算法能够提高推荐的准确率和召回率, 并降低 MAE 值。但是, 提出的算法在算法效率(如时间复杂度)上较传统算法并没有较大提高, 需要进一步优化。

参考文献

- [1] Xu Hai-ling, Wu Xiao, Li Xiao-dong, et al. Comparison Study of Internet Recommendation System[J]. Journal of Software, 2009, 20(2): 350-362 (in Chinese)
许海玲, 吴潇, 李晓东, 等. 互联网推荐系统比较研究[J]. 软件学报, 2009, 20(2): 350-362
- [2] Linden G, Smith B, York J. Amazon. com recommendations: Item-to-item collaborative filtering[J]. Internet Computing, IEEE, 2003, 7(1): 76-80
- [3] Sarwar B, Karypis G, Konstan J, et al. Item-based collaborative filtering recommendation algorithms[C]// Proceedings of the 10th international conference on World Wide Web. ACM, 2001: 285-295
- [4] Deng Ai-lin, Zhu Yang-yong, Shi Bo-le. A Collaborative Filtering Recommendation Algorithm Based on Item Rating Prediction[J]. Journal of Software, 2003, 14(9): 1621-1628 (in Chinese)
邓爱林, 朱扬勇, 施伯乐. 基于项目评分预测的协同过滤推荐算法[J]. 软件学报, 2003, 14(9): 1621-1628
- [5] Sun Guang-ming, Wang Shuo. Novel collaborative filtering algorithm based on interest degree of item[J]. Application Research of Computers, 2013, 30(12): 3618-3621 (in Chinese)
孙光明, 王硕. 基于项目兴趣度的协同过滤新算法[J]. 计算机应用研究, 2013, 30(12): 3618-3621
- [6] Wang Wei-ping, Wang Jin-hui. Hybrid Recommendation Method Based on Tag and Collaborative Filtering[J]. Computer Engineering, 2011, 37(14): 34-35 (in Chinese)
王卫平, 王金辉. 基于 Tag 和协同过滤的混合推荐方法[J]. 计算机工程, 2011, 37(14): 34-35
- [7] Xu Zhi-hong, Wang Bao-ying. Collaborative filtering algorithm based on item complex similarity[J]. Application Research of Computers, 2014, 31(2): 398-400 (in Chinese)
许智宏, 王宝莹. 基于项目综合相似度的协同过滤算法[J]. 计算机应用研究, 2014, 31(2): 398-400
- [8] Wang Bing-yi, Liu Yang, Nie Chang-xin. Personalized Recommendation Algorithm Based on Three-dimensional User Interest Modeling[J]. Computer Engineering, 2015, 41(1): 65-70 (in Chinese)
王冰怡, 刘杨, 聂长新. 基于用户兴趣三维建模的个性化推荐算法[J]. 计算机工程, 2015, 41(1): 65-70
- [9] Fang Jun-ming. Development and Challenge for Cognitive Psychology of Information process[J]. Psychological Science, 1998, 21(6): 481-484 (in Chinese)
方俊明. 信息加工认知心理学的发展和面临的挑战[J]. 心理科学, 1998, 21(6): 481-484
- [10] Lathia N, Hailes S, Capra L, et al. Temporal diversity in recommender systems[C]// Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2010: 210-217
- [11] <http://kddcup.yahoo.com/datasets.php#>
- [12] Zhao Liang, Hu Nai-jing, Zhang Shou-zhi. Algorithm Design for Personalization Recommendation System[J]. Journal of Computer Research and Development, 2002, 39(8): 986-991 (in Chinese)
赵亮, 胡乃静, 张守志. 个性化推荐算法设计[J]. 计算机研究与发展, 2002, 39(8): 986-991